Learning Environment. *Educational Technology & Society*, *15* (4), 27–38.

PEDALE – A Peer Education Diagnostic and Learning Environment

Johannes Konert[1], Kristina Richter[2], Florian Mehm[1], Stefan Göbel[1], Regina Bruder[2] and Ralf Steinmetz[1]

[1]Technische Universität Darmstadt, Multimedia Communication Lab, Rundeturmstraße 10, D-64283 Darmstadt, Germany // [2]Technische Universität Darmstadt, Department of Math Didactics, Schloßgartenstraße 7, D-64289 Darmstadt, Germany // johannes.konert@kom.tu-darmstadt.de // richter@mathematik.tu-darmstadt.de // florian.mehm@kom.tu-darmstadt.de // stefan.goebel@kom.tu-darmstadt.de // bruder@mathematik.tu-darmstadt.de // ralf.steinmetz@kom.tu-darmstadt.de

ABSTRACT

The basis for individual students' instructional support by teachers is an individual diagnosis of one's learning advances and difficulties. Even though sophisticated diagnostic tools exist, it remains an open question how diagnosis and learning can be merged into a consistent pedagogical method to support both teachers and students with feedback about the learning process. Aspects like the usage of open format questions, social inter-dependencies and heterogeneity, group-forming processes and the teachers supervision issue are taken into account. We propose a model for integration of peer assessment functionality for learning into a computer-based Adaptive Diagnostic Learning Environment to solve central problems of classroom diagnostic assessment, adaptive learning and knowledge transfer between peers in a classroom environment. The research approach focused on math learning scenarios for evaluation, but is expected to be applicable for other educational content as well. In the following paragraphs we identify several problems in classroom scenarios to be addressed, describe the model of the underlying approach, show the implementation details and explain the evaluation setup.

Keywords

Learner-centered Diagnostic Assessment, Technology-enhanced Learning, Knowledge Sharing

Introduction

Effectively supporting secondary school students in the classroom with appropriate learning material is a difficult task for a teacher. Despite the limited time for each student, instruction needs a proper diagnosis of each individual student's competencies and difficulties in order to make adequate didactic decisions for further instructional support.

Currently, tests used to assess students' understanding of certain topics are primarily paper-based (Howell, 2003). The results are aggregated by the teacher and afterwards feedback is provided to the students. These tests and diagnostic surveys have mainly been developed and accurately proofed from a psychometric point of view (Leighton/Gierl,2007; Bayrhuber er al., 2010). This means a very precise and narrow focus on valid measurement, but is incompatible with daily classroom instruction as their evaluation is too work-intensive for teachers and the tests stay isolated from the intended learning processes. Using software tools instead can support teachers in assessing the students' performance faster. Additional support for test adaption helps bridging the gap from diagnosis to instructional support.

Envisioned is a digital learning environment that allows the management of diagnostic and educational content and instructional feedback with the help of peer assessment and thereby affords bridging the gap between diagnosis and learning.

Related work

E-Learning environments that can serve diagnostic purposes are in the scope of different research interests. Intelligent Tutoring Systems try to keep track of the student's input and are predefined in a very elaborate way (Ritter et al., 2007). They are proved to be useful providing feedback in time of the problem solutions. In addition didactic research proved the advantages of methods like learning by teaching, collaborative learning and the benefits of social exchange for motivation (Gillies, 2004).

The impact of peer assessment has been studied earlier (Damon, 1984), highlighting the positive effects like raising interest for challenging tasks and fostering pro-social behavior. Beside a general interest of students in examining peer work (Stepanyan et al., 2009) better knowledge acquisition has been shown for computer-supported collaborative work as well (Mohammad et al., 2009). From the field of computer science the Social Network Analysis has carried out extensive research to investigate the phenomenon of knowledge sharing over weak ties between users not closely related (Petróczi et al., 2006; Granovetter, 1973; Fetter, 2009). Strangers are strong providers of help and knowledge without a directly expected reward (Constant et al., 1996).

Software Adaption to the progress of individual learners is especially investigated in the field of Serious Games for Learning (Bellotti et al., 2009; Tan et al., 2007) as this field combines the challenges of dynamic reactions to user behavior in the game as well as the learned skills. Therefore software maintains a player model to adapt to characteristics of decision behavior in games and updates the probabilities of learned skills in a learner model. These models' states parameterize specific issues like next scenes, challenges and information displayed to users.

Motivation – Technology-enhanced Diagnosis and Learning

Using software for diagnosis *and* learning still faces some conceptual and technical challenges:

Computer Diagnosis Problem

Processing and interpreting free text answers, drawings and different solutions of open format questions is still a challenge for computer systems. Advances in text and language processing are made continuously, especially if the context can be narrowed to a specific field. Nonetheless, the matching of semantic meaning in a student's reply to the desired answer remains as a research field. Thus diagnostic software tools widely use multiple-choice, gap text or sorting rather than open format test questions (Ritter et al., 2007). Unfortunately open format test questions are the most important ones for teachers from a diagnostic point of view as they reveal misconceptions or partial understanding of students (Prediger et al., 2008). We call this the *Computer Diagnosis Problem*.

Individual Group Assessment Problem

Diagnosis is usually conducted on an individual level. This prevents students from working collaboratively, sharing knowledge and giving hints. At the same time the benefits of group learning has been reported in many studies (Gillies, 2004; Klawe/Phillips, 1995). More precisely classroom research shows advantages for learning when feedback is given by peers as well rather than by teachers only (Gillies, 2004). Peer tutoring helps students to understand their misconceptions better, if they are explained by other students as they use the same language and share a common background for communication (Damon, 1984). We call the fact, that for individual diagnosis students need to be assessed individually, but for learning knowledge sharing in the peer group is favored the *Individual Group Assessment Problem*. It is desired to allow the knowledge sharing without risking precise individual students' assessment.

Peer Matching Problem

When students have to choose their peers in the classroom for a group work they usually feel obliged to choose their friends or peers of a similar proficiency level in the subject (Cohen, 1994). Both lead to a suboptimal, homogeneous group formation concerning instead of heterogeneous groups for optimal knowledge exchange and learning outcome for every group member. In secondary schools with classes that usually contain around 30 students a teacher has not the time to establish an optimal grouping for group work in pairs or triples as this would mean an intensive preparation to mix the students with different proficiency levels. Additionally learning styles should be taken into account for peer matching as it influences the perceived suitability of the group members and learning effects. Unfortunately students "tend to be rebellious if they are forced to work in groups that are not of their own choosing" (Mitchell et al., 2012). We call the fact that a mix of proficiency level is desired and learning styles should be considered for optimal knowledge exchange, but actually friendship and similar proficiency levels are matched the *Peer Matching Problem*.

Diagnosis Adaption Problem

The main goal of diagnosis is to provide a standardized and comparable result of individuals (OECD, 1999). A student's motivation for participation (i.e. using the tools provided) increases significantly, if the questions provided fit her individual skills and prevent situations of boredom or anxiety (Buchanan/Csikszentmihalyi, 1991). We call the fact that a suitable level of diagnostic task difficulty increases the motivation and performance of a student while diagnosis needs to be inter-individually comparable the *Diagnosis Adaption Problem*.

Teacher's Supervision Problem

The peer learning scenario where each student has his own pace and different tasks while sharing knowledge through feedback to each other by means of a computer-based environment is much more dynamic than traditional classroom setups. In order to keep control of the guidance and support of the class as a whole and each individual student at once, the teacher needs to have a tool at hand to supervise and influence the scenario in order to give the individual instructional support and have a diagnostic overview. The requirements concerning the teacher supervision can be summarized as the *Teacher's Supervision Problem*.

Scenario and Concept

We propose a system called PEDALE, as a novel approach addressing the above-mentioned problems by combining diagnosis and learning together with social networking principles for peer assessment and knowledge sharing between students. To the best of our knowledge no software with such an approach exists. In order to address the problems stated above, the system will use a carefully reviewed and empirically validated didactic model of competence development and diagnosis. Hence, PEDALE aims to be highly valuable for diagnosis (teacher's perspective) and understanding the own learning progress (students' perspective).

The proposed system will be used by teachers during classroom instruction to get a detailed diagnosis about their students' competencies. The students are instructed to use the software within a fixed time period (e.g. 40 minutes, depending on test configuration) to solve the diagnostic tasks, each student at an individual computer. During the time the students work with the software the teacher can monitor as well as participate in the process. With the help of a specific control panel that is activated if a teacher logs into the scenario the teacher can get an overview about the whole classes' progress as well as over certain events. It provides a filter-based search interface to see answers in the database by student or by task, with or without feedbacks (see fig. 4). The teacher can select a particular solution to be displayed like the feedback giving students see it. The teacher can simply look at the given feedbacks as well as give individual feedback to specific students himself. The control panel can slide up and down to overcome overlapping due to screen size restrictions.

Role Model

The users of PEDALE belong to two user groups: teachers and students in secondary schools. The teachers have the role of editing, changing and storing the scenario setups with the appropriate authoring software. In the player software they have a 'bird's eye view' over the scenario and can see which student has solved which tasks, given which feedback and so on (see below). The students are the second role. They open the configured scenario in their player software and solve the prepared tasks, give and receive feedback.

Authoring and Multi-Player Environment

Beside other application areas, the design of educational software faces the problem that the main experts (e.g. teachers) for the content used in the software are not programmers and vice versa. To decouple the dependencies during development a feasible approach is to provide authoring software for teachers to create content and configure the application behavior independently from programmers who otherwise would need to implement this. A second component is a player that displays the configured test interface and content to the students. The Authoring Tool will

be used for the setup of diagnostic tests and the input of test questions fitting the used diagnostic model. The corresponding player has to be capable of displaying the new interface elements and will adapt the test course.
By this approach we benefit from two key advantages:

1. the use of an authoring environment for teachers makes it easy to create class-specific e-learning content and can lead to better learning results (Mehm, 2010),
2. the use of a software-based player component provides a comfortable way of data retrieval for retrospect diagnostic purposes. Real-time results, quantitative and qualitative measures can be displayed in a specific teacher's view optimized for supervision, as well as in a student's view comprising his individual quantitative and qualitative measures.

A diagram of the software components with their key functionality and the data flow are displayed in fig. 1. The work with the software is arranged in three phases:

First, the *Assessment Setup* with teachers authoring, creating or selecting the desired test questions and setup the characteristics like duration, amount of peer assessment and the class setup (students).

Second, during the *Assessment* students load the configured test via their player software and work through the diagnostic assessment in the classroom (displayed as Student A). In the first phase of the assessment the students solve machine-analyzable tasks. On the base of these tasks a first diagnosis is generated automatically and returned to the students after they went through all the tasks of the first part. The second part of the assessment asks the students to evaluate solved problems regarding the correctness and the solution process. The answers to these solved problems are open test questions and are displayed to peer students (e.g. student B gets a solved problem of student A and vice versa). The solved problems are retrieved from the tool's data repository and the player decides which of the related solved problems matches best to be displayed. A Peer Matching Algorithm will be developed that takes into account students' current skill competence profiles and test performance.

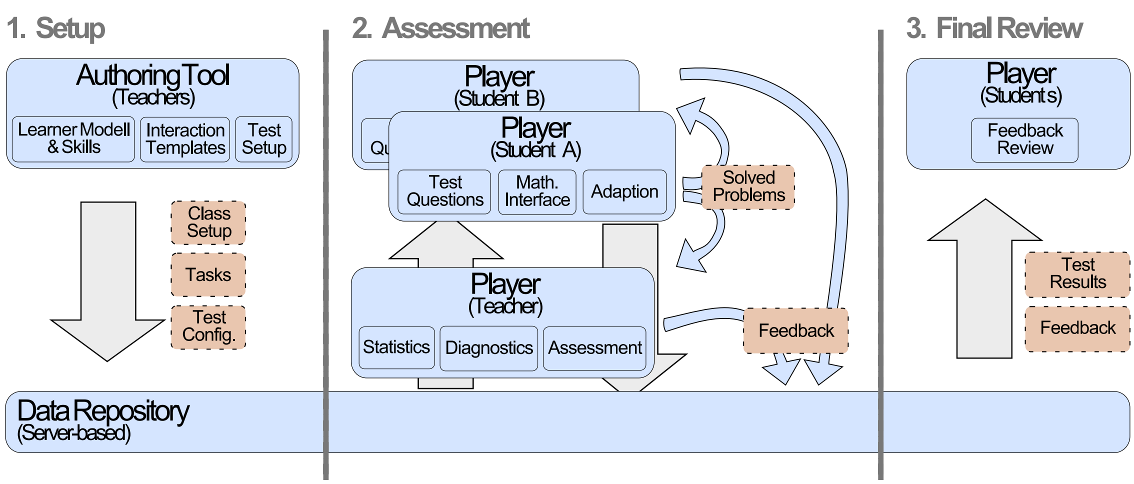

Figure 1: Phases (1-3) of diagnostics and learning with peer assessment

The given peer feedback is then stored for later review. Additionally, a second player version is provided to teachers for monitoring the students' progress and for final review of all solved problems.

In a final *Feedback* phase the students are provided with all their assessment results and peer feedback, as well as a feedback from the teacher.

Adaptive Diagnostic Model

As a sound diagnosis of the students' current state of knowledge is required for effective and individual learning and teaching, we use the diagnostic instrument developed in the 3-year project HEUREKO (Bayrhuber et al., 2010). Within the project a competence model for the mathematical domain "functions and graphs" for lower secondary level students was developed. Focus of the model is the heuristic use and change between the fundamental mathematical representations (numerical, graphic, symbolic, verbal) what can be considered as a significant competence of mathematical problem-solving and modelling (Bayrhuber et al., 2010). Theoretical didactic models of ability that have proven successful at a national as well as an international level were operationalized and empirically assessed in order to provide an empirically grounded instrument for diagnosis and instruction that can be applied to school practice. The Rasch analyses proved a four-dimensional model to be the best predictor. Furthermore, the separateness of these dimensions could be shown. Latent class analyses indicate that seven typical competence profiles can be identified empirically across the model dimensions (Bayrhuber et al., 2010). The resulting competency model here provides the basis for a diagnostic instrument for mathematical competencies in the domain "functions and graphs", while at the same time offering approaches to instructional support. The underlying model maps four dimensions of competencies on three levels of mastery. The first two levels comprise tasks that require a predefined input like multiple choice, decisions and numerical solutions, plotting points, intercepts or intersections. The highest level comprises open format replies like describing and reasoning.

We are transfering the paper-based tests about the understanding of mathematical functional dependencies into a software-represenation and provide an user-interface that allows students to choose between and produce verbal expressions, to draw graphs, to develop algebraic terms and to note and complete numerical representations.

About ¾ of the questions can be assessed automatically by the software as the solutions and results are definite. Variants of correct and incorrect students' solutions are taken from the results of the HEUREKO-Research Project. The questions that ask for open text input and the corresponding given answers (Solved Problems) are assessed by the teacher and peers while the student continues solving the next tasks. PEDALE uses the results to update the internal didactic learner model and select further questions accordingly. This adaptive diagnosis is possible without risking the comparability of the results due to the didactic model behind. The test questions are all categorized into several dimensions mapping exactly the tested competencies. Consequently the use of the appropriate didactic model avoids the *Diagnosis Adaption Problem* stated above. Still it can utilize the findings of flow theory research and adaptation. From the adaptive diagnosis approach we convey the first research question to be answered by evaluation:

RQ1: Does the Enhancement of diagnosis functionality with adaption enhance the learning outcome and acceptance by students?

Peer Assessment

The test itself will be organized into several parts, each containing questions for specific dimensions of the model. With the completion of one part of the underlying diagnostic model a student (Student B) is asked to review so called solved problems of this domain. These are questions that display the approach and/or solution of another student (Student A) and that ask student B:

- to decide whether or not the approach is correct and to rate the confidence of the given evaluation on a five-point-Likert-scale,
- give qualitative feedback on where things were done well, which mistakes can be identified or where insufficiencies were found,
- give hints and advice how the solution could be improved or solved alternatively and finally
- self-evaluate how helpful the given feedback might be for the addressed peer on a five-point-Likert-scale.

In order to give a constructive and helpful feedback each student has a feedback guide at his desk which contains guiding questions for writing a good feedback. The feedback guide is structured by what sort of solutions the students might find and differs between the given solution is 'correct', 'incorrect' and 'there's no solution'.

The peer assessment helps solving the *Computer Diagnosis Problem*. It enables us to provide open test questions with PEDALE and still get a reasonable assessment result. The learners' assessment of peer solutions is of great value for the learning process as it prompts the students to reflect a given solution and set it in relation to their own approach and knowledge. In doing so students are encountered with (a) real solutions and (b) approaches and mistakes of students with same social and learning background (Hilbert et al., 2008).

RQ2: Does assessing solutions of other students enhance learning? Do students gain a better awareness of their own knowledge about the assessed domain?

For peer assessment the effects of social networks have to be taken into account. Conflictive forces influence the student's motivation to invest time and energy in providing a good or average feedback to peers. Research in Social Network Analysis shows complex interdependencies between individuals in a social network. Studies show for settings in which people feel themselves as part of a common organizational team (like a school or class) a strong motivation to help each other with constructive and qualitative feedback (Constant, 1996). However, in a classroom environment a competitive situation and complex social interdependencies can exist. The influence on the peer feedback in this scenario remains an open research question. We investigate with one setup for evaluation, whether students provide more appropriate feedback to peer students when names are displayed or when the solution and feedbacks are displayed in an anonymous way. It is expected that students in general have the desire to see and comment other students' solutions as research for computer-supported collaborative learning environments indicates (Stepanyan et al., 2009). The proposed peer assessment setup is expected to support group learning aspects with knowledge sharing through feedback and to prevent the *Individual Group Assessment Problem*, because the students still carry out the test parts independently from each other.

RQ3: Which impact on acceptance of peer assessment has social anonymity? Is the perceived usefulness of giving and receiving feedback when the solutions and the feebdacks are anonymized?

Provision of Feedback

When the assessment time is over students are provided with a direct feedback. The tool returns an evaluation of the machine-analyzable questions as well as the feedback given by peers and the teacher. As Social Network Analyses indicate, the level of trust plays a major role for giving advise and critics (Petróczi et al., 2006; Golbeck, 2005). The transferability of effects of trust and closeness for classroom settings remains to be explored. As students share a more similar cultural background, language and interests with their peers as they do with the teacher, it is expected that feedback of other students is valued as a positive additional learning source. As the overall feedback is displayed after the test, it does not raise the *Individual Group Assessment Problem*.

RQ4:Are students more motivated to use the tool when they have the possibility to reflect other students' solutions and can give feedback?

Peer Matching Model

For each student the actual performance in the scenario (correct and incorrect solutions to tasks), the current math proficiency level (last math mark), gender and age are stored in the role model. In an extra questionnaire the learning style preferences are investigated and added to the model afterwards (see evaluation). As it remains uncertain which criteria should be considered to which degree for matching the peers for feedback provision and receiving, the model will store the mentioned parameters, but not use them for matching in our model so far. The authors expect to find indicators for optimal matching by analyzing statistically dependencies between the described criteria and the perceived usefulness of received feedback (rated by the students individually). Currently the model will be optimized to take criteria into account for distributing the matching randomly among all participating students that each student gives and receives a balanced amount of feedbacks. By matching the students automatically by computer-algorithms PEDALE helps solving the Peer Matching Problem as teachers do not need to match the students manually.

RQ5: Which influence have gender, math proficiency level and learning style on the perceived usefulness and acceptance of feedbacks by students?

Supervision

While the students are working with the software the teacher can monitor the classes' overall progress as well as individual student's solutions and feedbacks. Teachers can monitor the task solutions through a teacher supervision panel which allows them to filter the collected information according to their diagnostic or instructional interest. They can supervise all solutions to a specific task, all feedbacks to a specific solution or to a specific task, all solutions and feedbacks a specific students has submitted or received or watch the number of solved tasks in a general overview. Additional to monitoring the student's work, the teachers can intervene by writing feedbacks to a particular student's solution themselves or give hints how to solve the task when a student is stuck with a particular task. If desired teachers can intervene as well when an incorrect or inadequate feedback is given.

Implementation

Authoring and Multi-Player Environment

To provide a software tool for classroom learning the appropriateness for the specific classroom situation and the teacher acceptance depends on the configurability and content changeability. This must be easily achieved by domain experts (e.g. teachers) who normally do not have extensive programming skills.

We decided to build on two software components developed in our own research group, because they are easily extendable and proved their flexibility in several projects.

StoryTec (Mehm, 2010) allows teachers to set up the classroom characteristics and select, change or create the tasks and their order in the scenario setup without any programming skills necessary. More precisely *StoryTec* is based on the principles of Digital Storytelling. The flow of activities that later occur in the player is visualized to the author (e.g. the teacher) for editing by a graph of connected elements called scenes (see fig. 2). The appearance of scenes can be set up in a WYSIWYG-like editor. Beside video, sound, text-explanations and images, interactive elements like text-inputs, multiple-choice and handwriting support can be used for the scene design. Flow characteristics like time restrictions for single scenes or groups of scenes can be set. Teachers can configure which task scenes should be automatically assessed and specify the conditions. For task scenes that are not automatically assessed the teacher can configure the conditions for peer review (see III.D).

By manipulating the properties of scenes in several provided text fields, checkboxes and dropdowns the parameters are set to define whether a scene is an instructional one, a math task to be solved and stored to the database or a scene to request or display feedback. Teachers can easily arrange the scenes, connect them as well as create and group new tasks to be solved. The scenes can be cascaded to group elements and inherit properties from others. For the PEDALE scenarios the scene types for (task) *result storage*, (task) *containers*, *giving feedback* and *displaying feedback* have been added. Further details are omitted here and can be requested from the authors if the reader has interest in more technical details.

Based on our research we found the following setup of scenes for a scenario recommendable as a basis for individual adaption (see fig. 2):

1. solving two closed-format tasks on a comparably easy level, followed by automated diagnosis of the performance,
2. solving two open-format tasks and sending them to the system,
3. giving up to four times feedback to such open-format tasks' solutions of peers (depending on the time spent in the previous steps. The less time was spent in step 1 and 2, the more often feedback should be given),
4. reviewing received feedback,
5. re-editing formally not correctly solved tasks (or skipping in case of all correct),
6. equally to step 1. (solving two closed-format tasks)
7. equally to step 2. (solving two open-format tasks) and finally
8. a last review equally to step 4.

Beside the authoring tool *StoryTec* there is the player software *StoryPlay* (formerly known as *BatCave*) (Mehm, 2010a) for the students that loads all the data for the configured scenario and is capable of displaying the tasks, connecting with the database to read and write the task answers and finally controlling the flow of the scenario as configured. It has been extended to display the feedback requesting dialog and has a multi-user capability to request login-information and store as well as receive written solutions, handwriting notes and feedback elements to and from the database.

Both components, authoring tool and player, use the XML-based format for narrative game-based learning objects to exchange all dependencies and rules of the classroom scenario elements (Göbel et al., 2010). *StoryTec* as well as *StoryPlay* are flexibly extendible and proofed their validity as authoring and player software already for learning scenarios in the research field of Serious Games (Mehm, 2010a; Göbel et al., 2009).

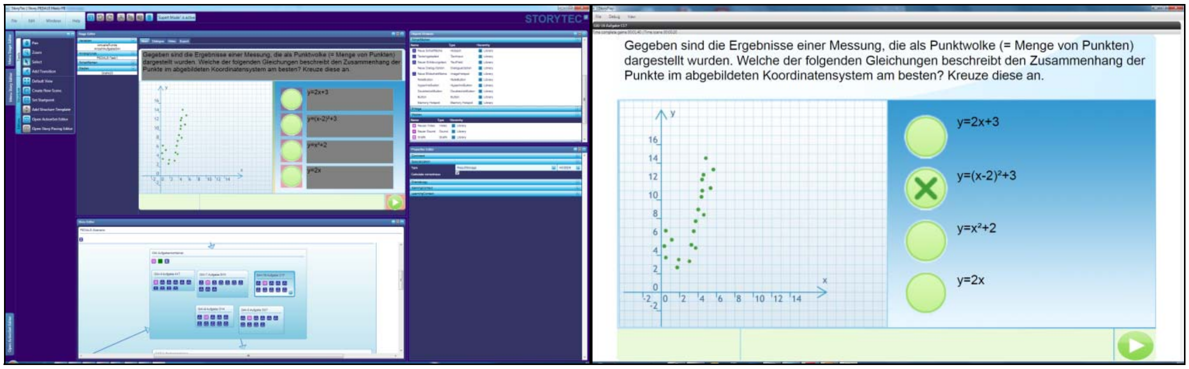

Figure 2: StoryTec (left) and StoryPlay (right) displaying the same math task scene (all texts in German)

Adaptive Diagnostic Model

For evaluation of the stated research questions we re-created several math tasks from the paper-based tests of the underlying diagnostic model as math task scenes in the authoring software *StoryTec*. This includes tasks with multiple-choice answers and open-format questions with the text-prompt to use the digital pen for algebraic, numeric or graphic approaches. As it is not the focus of the upcoming evaluation, not all dimensional dependencies are assessed and learned in the current setup. All mathematical tasks that are assessed here require a transformation from graphical to algebraic representations of functional dependencies. As respected in the proposed solution model of PEDALE it is afterwards desired to re-create further math tasks of differing representational changes as well and evaluate the adaption and selection of the tasks from different diagnostic model categories, too.

Peer Assessment

For better insight into the student's approaches, especially for the open-format tasks, PEDALE supports the use of digital pens, mouse and Microsoft stylus events. To allow for a handwriting and calculation that is as natural as possible, in our scenario students use a digital pen and write directly on regular paper. The pen movements are recorded and stored as an image. Several pages are possible for longer calculations. These images are then embedded in the respective task and are re-displayed to the peer students when students give feedback to each other, to the teacher during review of stored solutions and to the student himself when revising wrongly solved tasks.

When students are requested to give feedback to a peer's solution the best fitting candidate is selected from current database status. The selection is mainly based on the number of already received feedbacks to balance the knowledge exchange. If several candidates exist that have the same low number of received feedbacks, the candidates are further sorted binary by several subsequent criteria:

1. selecting a solution to assess, that is not of the same math task that has been assessed before (ensuring task variety),
2. selecting solutions of candidates, that have not already received feedback by the user (ensuring feedback sender variety),
3. selecting solutions of candidates, where the math task has not been solved by assessing user herself before (ensuring assessment variety for assesse).

Further criteria considered for later implementation depending on the evaluation results are gender, math grade and current scenario performance, learning style preferences, social network relations.

The assessing student is provided with the respective task and the peer's solution, i.e. the hand-written notes, and a feedback panel with structured feedback criteria (see fig. 3). Here the student assesses the correctness and completeness and gives constructive feedback. Additionally he is asked to rank his own certainty of giving feedback. Using the scale-based judgment the software can update the learner models of both assessor and the person assessed.

After the provision of qualitative feedback in the free-text field the student finally assesses the usefulness of his own feedback and then clicks the known play-button to jump to the next *StoryPlay* scene. The feedback and selections are stored to the database.

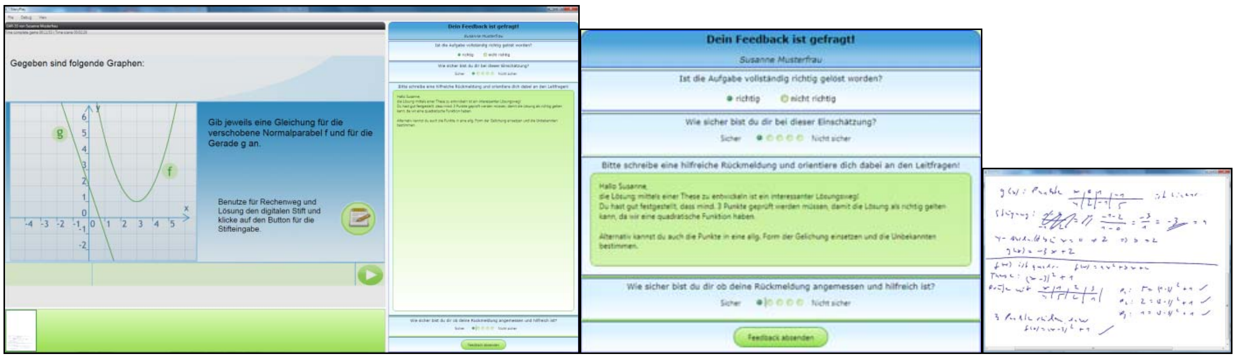

Figure 3: Peer Assessment screen with a solution of a peer student displayed on the left and the guiding questions to provide feedback on the right (enlarged as the image in the middle). The calculations of the peer can be opened in popup windows (see image on the right) (all texts in German)

Receiving provided Feedback

The qualitative feedback is displayed to the receiving student in a similar screen-layout (see fig. 4). When studying the received feedback students can freely switch between the several feedbacks by using an additional list or drop-down selection in order to compare different hints easily. Their own written notes can be re-opened for inspection. Additionally each feedback can be rated by the receiver on a five-point-Likert-scale. This rating can be taken into account for the evaluation to correlate the self-estimation of feedback-senders about their feedback usefulness and their self-assurance about the correctness and with the perceived usability.

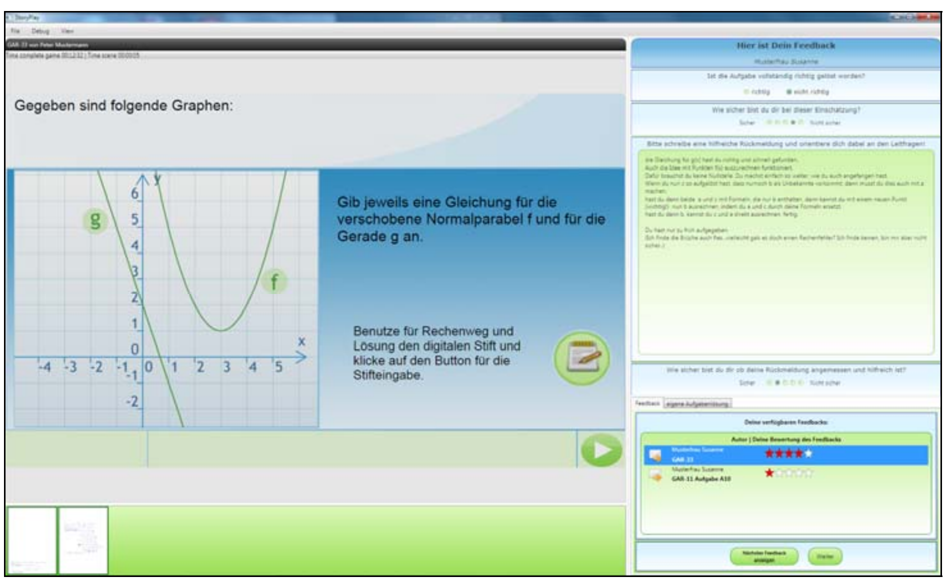

Figure 4: Example of the Student Feedback Receiving screen where the provided feedbacks are listed on the lower right, can be rated with stars and the own written solutions (left) can be re-opened in popup. On the right top the provided feedback of the peer (all texts in German)

Supervision

We enhanced the player software *StoryPlay* with a specific control panel that is activated if a teacher logs into the scenario. It provides a filter-based search interface to see answers in the database by student or by task, with or without feedbacks (see fig. 5). The teacher can select a particular solution from the list. It is displayed equally as to the feedback giving students. The teacher can look at the given feedbacks as well as provide individual feedback to specific students himself. The control panel can slide up and down to overcome overlapping issues due to screen size restrictions.

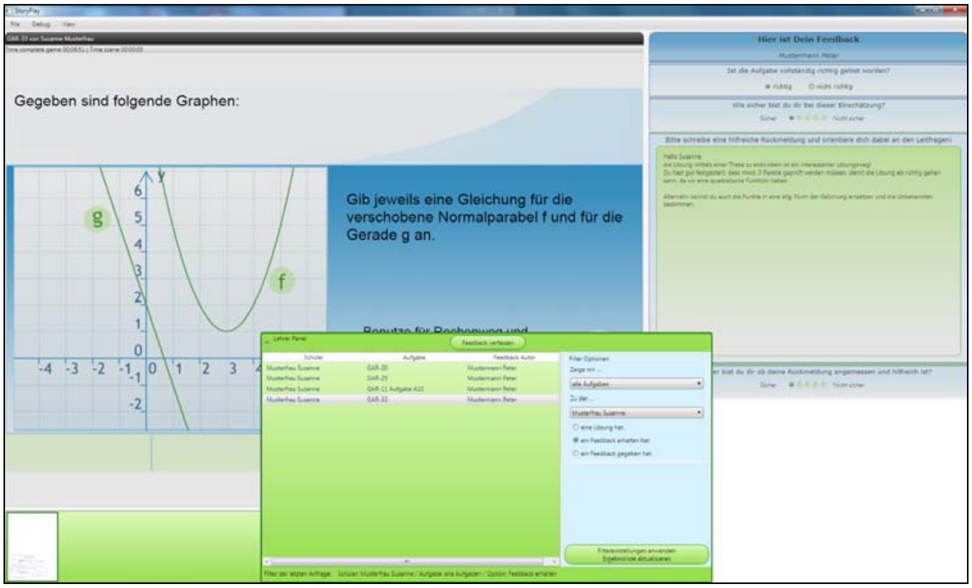

Figure 5: The additional control panel for teachers to list, filter and select items for review and the possibility to give feedback as well in an exra window opening by pressing the button (all texts in German)

Evaluation Approach

Before the main evaluation we conducted a pilot study with mathematics and teacher education experts in order to test the feasibility and acceptability of the software. The results involved the need of some minor design improvements as well as the suggestion of a teacher supervision panel and are already integrated into the concept introduced here. The main evaluation focusses on the technical feasibility and the validation of the electronic instrument at the beginning of the school year 2011/12 in seven 9th grade classes of three different secondary schools. The assessed competencies are expected to be available for the students but need to be revived and checked by the teacher in order to get a status quo and plan the next instructional unit. The evaluation of the computer-supported diagnostic instrument will last three weeks with a 1.5 hour diagnostic test each week in seven school classes in parallel. The tests for the classes will consist of the following setups, each in one class:

Setup α: A PEDALE-based test with no intermediate feedback function as a reference group for feedback and social ties. (The reference group gives and receives feedback only at the end of the course after all the tasks solving is over.) The tasks are solved without any name recorded with the task solution. In short this setup variant is called “no feedback, anonymously”.

Setup β: A PEDALE-based test with no intermediate feedback function but name recording with the task solution as a reference group for feedback. In short: “no feedback, namely”.

Setup γ: A PEDALE-based test with intermediate feedback to and from peers without displaying names in the task solutions and the feedbacks as an indicator of the influence by feedback functionality. In short: “feedback, anonymously”.

Setup δ: A PEDALE-based test with intermediate feedback to and from peers with displaying names in the task solutions and the feedbacks as an indicator of the influence by social ties between students and anticipated competition. In short: “feedback, namely”.

Future Work

Due to the former research it is expected to find evidence for RQ1, RQ2 and RQ4 to support the core concept of the stated approach of combining diagnosis with peer assessment for learning. Competitiveness and complex social interdependencies between individuals in school class has certainly to be taken into account as a factor for peer assessment. However because the field evaluation described above cannot affect any marks of students and covers the mathematical content of the previous school year, students might not consider competitiveness during the peer assessment. Social interdependencies however might still be relevant in the evaluated scenario. This situation might result in no significant proof for RQ3. If there is evidence supporting RQ3, considering this aspect in further projects seems to be reasonable. Last, if beside RQ4 indications for RQ5 can be found, we will further intensify our investigation of social interdependency factors and peer matching criteria.

Technical issues for future work include a widening towards the inclusion of game elements in order to increase motivation and flow experiences. Additionally the creation of a web-based solution is considered that can be used by students not only in the classroom, but also accompanying homework to assess peers’ solutions, receive feedback and develop knowledge together in a *Social Adapting Diagnostic and Learning Environment*.

Conclusion

In this paper some crucial challenges of everyday classroom instruction have been described which affect traditional as well as technology-enhanced teaching. Although some problems remain open it has been shown how far digital learning environments could support those central processes of diagnosis and learning through knowledge exchange among peers. Hereby, the integration of the social network in the classroom seems to be a vital element of classroom learning that needs to be considered in digital environments as well. In our evaluation we focus on critical design challenges and analyze the benefits and potential that such a learning software has for teaching and learning.

Acknowledgment

The research and development introduced in this work was partly funded by the German Research Foundation (DFG-GK-1223).

References

Howell, S.L. (1995): E-Learning and paper testing: Why the Gap? *Educause Quarterly*, 26(4), 8–11.

Leighton, J. & Gierl, M. (2007): Cognitive diagnostic assessment for education. Cambridge: Cambridge University Press, 371.

Bayrhuber, M., Leuders, T., Bruder, R. & Wirtz, M. (2010): pedocs - Repräsentationswechsel beim Umgang mit Funktionen – Identifikation von Kompetenzprofilen auf der Basis eines Kompetenzstrukturmodells. Projekt HEUREKO. In: Klieme, E., Leutner, D. & Kenk, M. (Eds.) *Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes*, Weinheim: BELTZ Pädagogik, 28-39.

Ritter, S., Anderson, J.R., Koedinger, K.R. & Corbett, A. (2007): Cognitive Tutor: Applied research in mathematics education. *Psychonomic Bulletin & Review*, 14(2), 249-55.

Gillies, R. (2004): The effects of cooperative learning on junior high school students during small group learning. *Learning and Instruction*, 14(2), 197-213.

Damon, W. (1984): Peer education: The untapped potential. *Journal of Applied Developmental Psychology*, 5(4), 331-343.

Stepanyan, K., Mather, R., Jones, H. & Lusuardi, C. (2009): Student engagement with peer assessment: A review of pedagogical design and technologies. In: *Advances in Web Based Learning–ICWL 2009*, 367–375.

Mohammad, A. L. S. , Guetl, C. & Kappe, F. (2009): PASS: Peer-ASSessment approach for modern learning settings. In: *Advances in Web Based Learning-ICWL 2009: 8th International Conference, Aachen, Germany, August 19-21, 2009, Proceedings*, 2009, no. Iicm, 44.

Petróczi, A., Nepusz, T. & Bazsó, F. (2006): Measuring tie-strength in virtual social networks. *Connections*, 27(2), 39–52.

Granovetter, M.S. (1973): The strength-of-weak-ties perspective on creativity: a comprehensive examination and extension. *The American Journal of Sociology*, 78(6), 1360-1380.

Fetter, S., Berlanga, A. & Sloep, P. (2009): Enhancing the social capital of learning communities by using an ad hoc transient communities service. In: *Advances in Web Based Learning–ICWL 2009*, 150–157.

Constant, D., Sproull, L. & Kiesler, S. (1996): The kindness of strangers: The usefulness of electronic weak ties for technical advice. *Organization Science*, 7(2), 119-135.

Bellotti, F., Berta, R., De Gloria, A. & Primavera, L. (2009): Adaptive experience engine for serious games. *IEEE Transactions on Computational Intelligence and AI in Games*, 1(4), 264–280.

Tan, P.-H., Ling, S.-W. & Ting, C.-Y. (2007): Adaptive digital game-based learning framework. In: *Proceedings of the 2nd International Conference on Digital Interactive Media in Entertainment and Arts - DIMEA ’07*, 142.

Prediger, S., Selter, C. & Dortmund, U. (2008): Diagnose als Grundlage für individuelle Förderung im Mathematikunterricht. *Schule NRW*, 6(3), 113-116.

Klawe, M. & Phillips, E. (1995): A classroom study : Electronic games engage children as researchers. In: *The 1st International Conference on Computer Support for Collaborative Learning*, 209–213.

Cohen, E. G. (1994): *Designing groupwork: strategies for the heterogeneous classroom*, 2nd ed., Teachers College Press, p. 202.

Mitchell, S. N. , Reilly, R., Bramwell, F. G. & Lilly, F. (2012): Friendship and choosing groupmates: Preferences for teacher-selected vs. student-selected groupings in high school science classes. *Journal of Instructional Psychology*, 31(1), 1-6.

OECD (1999): Measuring student knowledge and skills: a new framework for assessment.

Buchanan, R. & Csikszentmihalyi, M. (1991): Flow: The psychology of optimal experience. *Design Issues*, 8(1), 80.

Chen, J. (2007): Flow in games (and everything else). *Communications of the ACM*, 50(4), 31-34.

Mehm, F. (2010a): BatCave: A testing and evaluation platform for digital educational games, In: *Proceedings of the 3rd European Conference on Games Based Learning,* Copenhagen, 251-260.

Hilbert, T., Renkl, A., Schworm, S., Kessler, S. & Reiss, K. (2008): Learning to teach with worked-out examples: a computer-based learning environment for teachers. *Journal of Computer Assisted Learning,* 24(4), 316–332.

Golbeck, J. A. (2005): Computing and applying trust in web-based social networks. University of Maryland.

Mehm, F. (2010): Authoring serious games. In: *Conference on the Foundations of Digital Games*, 271-273.

Göbel, S, Wendel, V., Ritter, C. & Steinmetz, R. (2010): Personalized, adaptive digital educational games using narrative game-based learning objects. *Entertainment for Education. Digital Techniques and Systems*, 438–445.

Göbel, S., Mehm, F., Radke, S. & Steinmetz, R. (2009): 80days: Adaptive digital storytelling for digital educational games. In: *Proceedings of the 2nd International Workshop on Story-Telling and Educational Games (STEG’09)*.